Batch variation between branchial cell cultures: An analysis of variance

Hansen, Heinz Johs. Max; Grosell, M.; Kristensen, L.

Publication date:
2003

Document Version
Publisher's PDF, also known as Version of record

Risø–R–1384(EN)

Batch Variation between Branchial Cell Cultures: An Analysis of Variance

Heinz J. M. Hansen, Martin Grosell &
Leif Kristensen

Risø National Laboratory, Roskilde, Denmark
January 2003

Abstract We present in detail how a statistical analysis of variance (ANOVA) is used to sort out the effect of an unexpected batch-to-batch variation between cell cultures. Two separate cultures of rainbow trout branchial cells were grown on permeable filter supports ("inserts"). They were supposed to be simple duplicates for testing the effect of two induced factors—apical or basolateral addition of radioactive precursors and different apical media–on the incorporation of ^{14}C-acetate and ^{32}P-phosphate into tissue lipids. Unfortunately, they did not altogether give the same result. By accepting this fact and introducing the difference between batches as one of the factors in an expanded three-dimensional ANOVA, we were able to overcome an otherwise crucial lack of sufficiently reproducible duplicate values. We could thereby show that the effect of changing the apical medium was much more marked when the radioactive lipid precursors were added on the apical, rather than on the basolateral, side. The insert cell cultures were obviously polarized. We argue that it is not reasonable to reject troublesome experimental results, when we do not know a priori that something went wrong. The ANOVA is a very useful statistical tool that can be utilized by any experimentalist who cares to make the effort.

ISBN 87-550-3163-3; 87-550-3161-7(internet);
ISSN 0106–2840

Print: Pitney Bowes Management Services Denmark 2003

Contents

1 Introduction

This investigation regards two issues. The first is one of the most common dilemmas in observational work in situations where there is not full control of the of experimental conditions. The second is an example of such a type of experiment.

1.1 Lack of Reproducibility

Imagine that you are working with a series of experiments based on batches of cell cultures (or any other group of biological cohorts). So far you have been getting reproducible results, but then something goes wrong: Two batches, supposed to be simple duplicates for testing a new variable, turn out differently. Your standard deviation gets out of hand and does not enable you to confirm the significance of an apparent effect. What do you do? Obviously, you repeat the experiment.

But what if you cannot just repeat it? What if it represents the last part of an important work which you are not in a position to continue as before? It would be a shame to give up. In fact, the unexpected extra variation might in itself supply valuable information.

Biologists and many other experimental scientists are working with systems which are often not under complete experimental control: The systems may be subject to unknown batch-to-batch variations or year-to-year cohort variations. The scientists are under strong obligation to rely on formal statistical methods to test the significance of observed differences. The so-called Student's t-test, which tests simple differences between two groups, has gradually been accepted by most experimentalists as the use of pocket calculators and computers has become more widespread. The corresponding, somewhat more sophisticated, *analysis of variance* (ANOVA) is less popular, even though it can also be carried out quite easily with standard statistical computer software. People are generally in doubt as to what an ANOVA actually means. It is often regarded as just a touch of witchcraft, used to appease 'difficult' reviewers. This, however, is not fair to its merits.

In the following we wish to present in detail how an ANOVA was used to sort out the effect of an unexpected batch-to-batch variation between cell cultures. By accepting the facts and introducing the observed extra variation as one of the factors in an expanded three-dimensional ANOVA, we were able to overcome an otherwise crucial lack of sufficiently reproducible duplicate values. Our aim is to show how this could be accomplished in a simple manner.

1.2 Lipid Metabolism in Branchial Cell Cultures

The example presented in the following is part of an extensive study (Hansen et al. 1992, Hansen et al. 2002) regarding the role of membrane lipids in stabilizing protein functions.

Branchial cell cultures have been used to mimic the regulation of ion transport in fish gills (Hansen et al. 2002). Wood & Pärt (1997) developed a method (Fig. 1) for growing branchial epithelia from rainbow trout (*Oncorhynchus mykiss*) gills on permeable filter supports ("inserts"). These epithelia can be exposed to different media on their apical and basolateral surfaces and, indeed, survive apical freshwater (FW) exposure (pseudo in vivo conditions) for up to 48 hours (Gilmour et al. 1998).

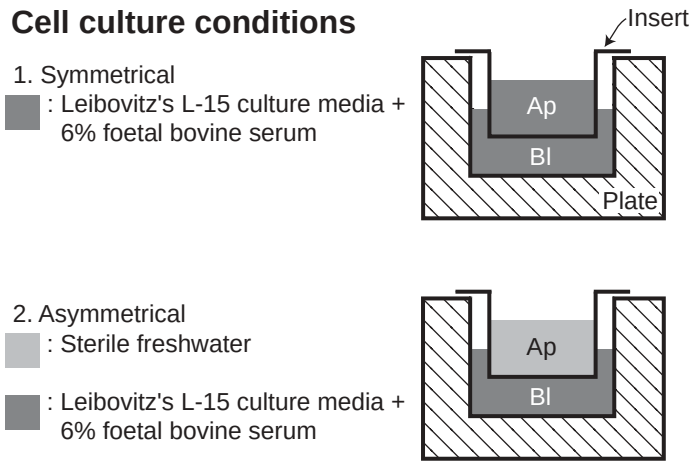

Figure 1. Coherent branchial cell cultures are grown on the permeable bottom of the insert cups, which are placed inside the plate wells—under symmetrical and asymmetrical conditions. Ap: apical medium. Bl: basolateral medium.

Cell-culture lipid metabolism was studied using ^{14}C-acetate and ^{32}P-phosphate as precursors (Hansen et al. 2002). ^{14}C-acetate labels the fatty acid moiety while ^{32}P-phosphate labels the polar head groups of membrane phospholipids.

The investigation followed a factorial experimental design. In the first experimental series the apical medium was varied between Leibovitz growth medium (L15) and FW, with L15 always on the basolateral side (Fig. 1). The radio-labeled precursors were added on the apical side. Reproducible results showed tissue phospholipids dominated by ^{32}P-phosphatidylethanolamine with FW on the apical side (Hansen et al. 2002), and the highest total incorporation of both precursors with L15 on the apical side.

The next step was to repeat the series with the radio-labeled precursors on both the apical and the basolateral side. We also wanted to test one more apical medium, which mimicked the osmolality of L15. This was accomplished by adding mannitol (M) to FW. Thus, the apical media were FW, FW+M, and L15. The basolateral medium was always L15.

2 Materials and Methods

Below we present the experimental setup and the method of data analysis. Figure 1 illustrates the former.

2.1 Incubation with Radio-Labeled Precursors

The experiments were carried out with two cell cultures (batches I and II) from different fishes on different days. ^{32}P-phosphate was added to the incubation media at 50 μCi/ml (1.5 or 2.0 ml in all) and incorporated into tissue lipids down to 0.05 nCi/insert, i.e. a dilution factor of about 10^{-6}. Consequently, the ^{32}P-phospholipids had to be carefully isolated by means of thin-layer chromatography before counting. It took three weeks to scan each of the three ^{32}P-incorporation

patterns at the minimum incorporation level. In view of the rather short ^{32}P half-life of only 14.3 d, we could afford just one duplication within each batch. We found that otherwise we would loose too much activity by radioactive decay before we reached the last counting period. ^{14}C-acetate was added (together with ^{32}P-Phosphate) at 20 μCi/ml and incorporated down to 1 nCi/insert. In view of the long half-life of 5570 y, the counting of ^{14}C was not a major problem. The radio-labeled lipids were assayed by thin-layer chromatography, using a special low-background radio-chromatogram scanner (Hansen 1975, Bøtter-Jensen et al. 1977).

2.2 Analysis of Variance

In the following we present a short introduction to an analysis of variance. Our aim is to show how an ANOVA works. We need to know how to interpret the various terms in the equations. The reader is referred to Hald (1952*b*), Bailey (1958), and Dytham (1999) for a comprehensive presentation.

While the Student's t-test considers the difference between *two* mean values in relation to the corresponding standard deviation and tests whether or not this difference could be explained by a random statistical variation within the limits of the observed standard deviation value, the analysis of variance estimates the components of variation due to a *series* of systematic and random causes. Our basis is the factorial experiment. Here we have a series of induced conditions or *factors* that are varied at various qualitative or quantitative *levels*. We wish to know whether these factors actually mean something, whether their effects are significant. Just as in the Student's t-test, we have the alternative option that an observed variation could be a random phenomenon. We end up by estimating the probability P that the alternative option, i.e. the random option, is true.

In an ANOVA we collect all the data in one bag without first considering the reasons for their mutual deviations. Then these data are sorted according to the factors we want to investigate, and it is assessed whether or not various ratios of variation differ from unity in a statistically significant manner. The basic assumption is the addition theorem, which states that the variances, not the standard deviations, are additive: if a and b are two statistically independent variables with the standard deviations s_a and s_b then the standard deviation of $a + b$ is $s_{a+b} = \sqrt{s_a^2 + s_b^2}$. The theorem can be generalized to state that the variation within a group of observations is best expressed as the *Sum of Squares of individual Deviations from the mean* (SSD). An ANOVA thus determines the total SSD—i.e. the sum of the squares of the deviations between each measured value x and a common total mean $\overline{x}$, thus $\mathrm{SSD}_{\mathrm{total}} = \sum(x - \overline{x})^2$—and partitions it systematically into a sum of individually specified SSD's in accordance with the experimental plan. The effect of an experimental factor is assessed by the extra variation, in terms of extra SSD, that it causes.

We consider a factorial experiment with two main factors A and B varied at k and m levels, respectively, and list the outcome in a two-dimensional scheme (matrix), where all the values in the same row refer to the same level of say the factor A, while all the values in the same column refer to the same level of B. We calculate the set means, row means, column means, and total mean. For illustration, Table 1 shows an example of the various means for $k = 2$ levels of A, $m = 3$ levels of B, and $n = 2$ observations within each set.

Table 1. Schematic presentation of observations with two systematic factors A *and* B *and with* $n = 2$ *observations within each set. This means that each element in the* $k \times m = 2 \times 3$ *matrix consists of* $n = 2$ *subelements.*

A \ B	$m=3$: 1	2	3	row means
1	x_{111}, x_{112}	x_{121}, x_{122}	x_{131}, x_{132}	$\overline{x}_{1..}$
$k=2$	$\overline{x}_{11.}$	$\overline{x}_{12.}$	$\overline{x}_{13.}$	
2	x_{211}, x_{212}	x_{221}, x_{222}	x_{231}, x_{232}	$\overline{x}_{2..}$
	$\overline{x}_{21.}$	$\overline{x}_{22.}$	$\overline{x}_{23.}$	
column means	$\overline{x}_{.1.}$	$\overline{x}_{.2.}$	$\overline{x}_{.3.}$	total mean $\overline{x}_{...}$

In this case we have

$$\begin{aligned}\mathrm{SSD}_{\mathrm{total}} &= (x_{111}-\overline{x}_{...})^2+(x_{112}-\overline{x}_{...})^2+(x_{121}-\overline{x}_{...})^2+(x_{122}-\overline{x}_{...})^2\\ &\quad+(x_{131}-\overline{x}_{...})^2+(x_{132}-\overline{x}_{...})^2+(x_{211}-\overline{x}_{...})^2+(x_{212}-\overline{x}_{...})^2\\ &\quad+(x_{221}-\overline{x}_{...})^2+(x_{222}-\overline{x}_{...})^2+(x_{231}-\overline{x}_{...})^2+(x_{232}-\overline{x}_{...})^2\,.\end{aligned}$$

As a measure of the effect of the two main factors A and B, respectively, we introduce the variations $\mathrm{SSD_A}$ and $\mathrm{SSD_B}$. $\mathrm{SSD_A}$ represents the deviations between the row means and the total mean $\overline{x}_{...}$, while $\mathrm{SSD_B}$ represents the deviations between the column means and the total $\overline{x}_{...}$. Thus

$$\mathrm{SSD_A} = n \times m \times \left[(\overline{x}_{1..}-\overline{x}_{...})^2+(\overline{x}_{2..}-\overline{x}_{...})^2\right]$$

and

$$\mathrm{SSD_B} = n \times k \times \left[(\overline{x}_{.1.}-\overline{x}_{...})^2+(\overline{x}_{.2.}-\overline{x}_{...})^2+(\overline{x}_{.3.}-\overline{x}_{...})^2\right].$$

There is also the effect of an *interaction* AB. $\mathrm{SSD_{AB}}$ covers the variation, besides the effect of the two factors A and B alone, due to any interdependence between them. It is made up of the deviations between the individual set means and the total mean, which cannot be accounted for by just regarding the corresponding row and column means. By this definition we have

$$\begin{aligned}\mathrm{SSD_{AB}} &= n \times \left[(x_{11.}-\overline{x}_{1..}-\overline{x}_{.1.}+\overline{x}_{...})^2+(x_{12.}-\overline{x}_{1..}-\overline{x}_{.2.}+\overline{x}_{...})^2\right.\\ &\quad+(x_{13.}-\overline{x}_{1..}-\overline{x}_{.3.}+\overline{x}_{...})^2+(x_{21.}-\overline{x}_{2..}-\overline{x}_{.1.}+\overline{x}_{...})^2\\ &\quad\left.+(x_{22.}-\overline{x}_{2..}-\overline{x}_{.2.}+\overline{x}_{...})^2+(x_{23.}-\overline{x}_{2..}-\overline{x}_{.3.}+\overline{x}_{...})^2\right]\end{aligned}$$

Finally, there is the random or *residual* variation within sets $\mathrm{SSD_0}$.

The basis for a partitioning is the following identity, with the terms presented above,

$$\mathrm{SSD_{total}} = \mathrm{SSD_A}+\mathrm{SSD_B}+\mathrm{SSD_{AB}}+\mathrm{SSD_0}.$$

The next step in an ANOVA is calculating a series of "standard" variances s^2, i.e. the squares of actual standard deviations. The generic expression for such a quantity is

$$s^2 = \frac{\mathrm{SSD}}{\mathrm{Df}},$$

where Df is the so-called number of degrees of freedom or just degrees of freedom. Df is connected with the number of experiments. It is a measure of the reliability of an observed variation in the sense that the confidence that can be attributed to a calculated variance is larger, the larger the value of Df.

All variances are mutually comparable. They represent, in contrast to SSD, a common level of variability, independent of experimental volume. The various calculated values of s^2 make up the essence of an ANOVA. They are compared with each other by forming the so-called F-ratios, i.e. s_1^2/s_2^2. The idea is to test whether or not these ratios differ from unity in a statistically significant manner.

The term degrees of freedom is likewise known from the Student's t-test, where the t-distribution table has Df as a variable. The larger the value of Df, the smaller the minimum t-ratio at a given probability P of making a mistake and accepting a random difference as true. If we, as an example, want to make sure that $P < 0.05$, then the t-ratio has to be larger than 4.30 at Df = 2, but only larger than 2.57 at Df = 5 (Hald 1952*a*). Similar considerations apply in an ANOVA, where we regard F-ratios between variances. The larger the value of Df pertaining to s^2, the smaller the the minimum F-ratio at a given level of significance.

We also know, from calculations in connection with the Student's t-test, that if we just repeat the same measurement n times, then $\mathrm{Df} = n - 1$. This means that Df = 0 when $n = 1$, corresponding to the fact that we cannot calculate a variance from only one observation. When $n = 2$ we can start calculating s^2 with Df = 1. We call the variance of n simple repetitions $s_0^2 = \mathrm{SSD}_0/(n-1)$, equivalent to the random residual variation SSD_0 within sets.

Similar conditions regarding Df apply in an ANOVA. If the factor A is varied at k levels then $\mathrm{Df_A} = k - 1$, just as B at m levels gives $\mathrm{Df_B} = m - 1$. It is not straightforward to derive an expression for $\mathrm{Df_{AB}}$. Here we will just state the result $\mathrm{Df_{AB}} = (k-1) \times (m-1)$ (Hald 1952*b*).

Apart from the random component s_0^2 of the calculated variances, the ANOVA also introduces the corresponding main factor variances s_A^2 and s_B^2, together with the interaction variance s_AB^2. These are equivalent to $\mathrm{SSD_A}$, $\mathrm{SSD_B}$, and $\mathrm{SSD_{AB}}$, respectively, but more specific in that, e.g., s_A^2 just covers the effect of the factor A alone, while $\mathrm{SSD_A}$ covers the effect of the factor A in general.

A further partitioning of the standard variances with the systematic experimental factors A and B, and the levels k, m, and n becomes (Hald 1952*b*)

$$\begin{aligned}
s_4^2 &\equiv \frac{\mathrm{SSD_A}}{k-1} = n\,m\,s_\mathrm{A}^2 + n\,s_\mathrm{AB}^2 + s_0^2, \\
s_3^2 &\equiv \frac{\mathrm{SSD_B}}{m-1} = n\,k\,s_\mathrm{B}^2 + n\,s_\mathrm{AB}^2 + s_0^2, \\
s_2^2 &\equiv \frac{\mathrm{SSD_{AB}}}{(k-1)\times(m-1)} = n\,s_\mathrm{AB}^2 + s_0^2, \\
s_1^2 &\equiv \frac{\mathrm{SSD}_0}{k\times m\times(n-1)} = s_0^2.
\end{aligned}$$

We note that the random residual variance s_0^2 occurs in all four equations. Furthermore, we note that s_2^2 is a part of both s_3^2 and s_4^2. This is the crux of the issue in the present use of an ANOVA, namely to sort out the effect of an unexpected extra variation. If the F-ratio between variances, $\mathrm{F} = s_2^2/s_1^2$, say, does *not* differ significantly from unity, the implication is that s_{AB}^2 can be assumed zero and that s_2^2 can be used, with its extra $(k-1)\times(m-1)$ degrees of freedom, as an additional measure of $s_1^2 = s_0^2$. In other words, when there is no true interaction AB, then $(k-1)\times(m-1)$ out of the $k \times m$ set means can be deduced from the others and thereby, when we actually measure them, be regarded as extra duplicate results. This strengthens the power of further F-tests, where the statistical significance of s_3^2 and s_4^2 can be tested against a combined

$$s_0^2 = \frac{\mathrm{SSD}_0 + \mathrm{SSD}_{\mathrm{AB}}}{k \times m \times (n-1) + (k-1) \times (m-1)}.$$

The method will be discussed in the following example, where it will also be applied to a three-dimensional ANOVA, i.e. in a case where there is one more factor than discussed above.

3 Results

The total incorporation results are shown in Fig. 2.

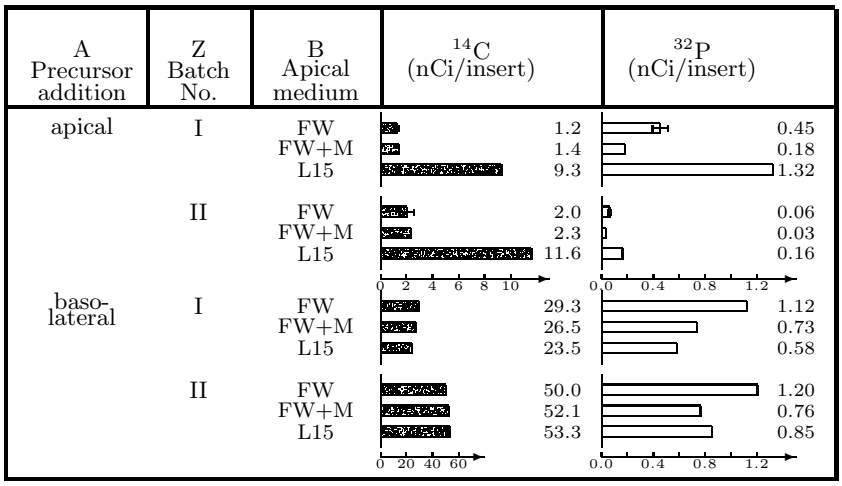

Figure 2. Trout branchial cultures, single seeded. Batch I and batch II from different fishes on different days. Incorporation of ^{14}C*-acetate and* ^{32}P*-phosphate into tissue lipids after incubation for 2 h. The observations are presented as horizontal bars as well as by numbers to the right. Note the different scales.*

3.1 Variation between Batches

It appears directly that the effect of changing the apical medium (the factor designated B) also depends on to which side of the culture the precursors have been added (the factor designated A). It follows that there must be an interaction AB, and that this seems to apply to both precursors, as well to the incorporation of

^{14}C-acetate as to the incorporation of ^{32}P-phosphate. However, there is a major problem: *The results from the two batches differ by as much as a factor of eight when we regard the total incorporation into tissue lipids of* ^{32}P*-phosphate added on the apical side.* This applies to all three apical media. Thus, the apparent interaction AB can only be confirmed as far as the ^{14}C-labeled lipids are concerned. It cannot be confirmed as far as the ^{32}P-labeled lipids are concerned, unless we disregard the apical incorporations in either batch I or batch II. This, however, we cannot just do. One should not ignore genuine experimental results, except when knowing, a priori, that they are wrong.

3.2 Incorporation of ^{14}C-Acetate

In the following a two-dimensional ANOVA is presented in Table 2. It is based on the logarithms of the ^{14}C results in both batches, which are regarded as simple duplicates. The extra two duplicate values within batches (with apical addition of precursor and apical FW) are included as means.

Table 2. ANOVA of ^{14}C*-incorporation from Fig. 2. Logarithmic values.*

i	Variation	SSD	Df	s_i^2	F-value	Fractile
4	A: Between precursor additions	3.499	1	3.499	$\frac{3.499}{0.207} = 17$	97.5%=12 ($P < 0.025$)
3	B: Between apical media	0.377	2	0.188	$\frac{0.188}{0.226} = 1$	
2	Interaction AB	0.452	2	0.226	$\frac{0.226}{0.031} = 7.3$	97.5%=7.3 ($P < 0.03$)
1	Residual (within sets)	0.187	6	0.031		
	Total	4.515	11			
	AB + B	0.829	4	0.207		

But why are we using the logarithms? In the analysis of variance we assume the same random variability, represented by $\pm s_0$, throughout. Further, the effect of each factor is considered additive, i.e. as adding or subtracting various absolute amounts to the result. Mean values are calculated as *arithmetic means*. However, biological effects are more often "multiplicative" than additive, i.e. they enhance or reduce the outcome by "scaling factors". The equivalent mean values are *geometric means*. Similarly, the standard deviations are proportional to the means. Taking the logarithm of each measured value amends all these difficulties since it transforms relative deviations into absolute deviations (and—at the same time—geometric means into arithmetic means). There is nothing unnatural in doing so. We are looking for significant differences and these may just as well show up when studying functions of the results as when we deal with the results themselves. Actually, statistical computer software always has an analysis of the logarithmic values as an option.

The computer is patient, so whether or not one should use the logarithmic values can be decided by utilizing both options and choosing the result that looks most

reasonable. In the present case it can also be seen directly. Figure 2 shows that the difference between the ^{14}C-values in batch II and batch I, here regarded as due to a random variation, is that those in batch II are about twice as high as those in batch I. Similarly, we see that the apical ^{32}P- values in batch I are about 8 times those in batch II, irrespective of the type of incubation medium. This makes us chose the logarithmic option.

Table 2 shows two clearly significant effects:

1. When tested against the residual variation—s_0^2 with 6 degrees of freedom—the interaction AB turns out to be significant (Hald 1952*a*)at the level $P < 0.03$; i.e. there is a less than 3% probability that the observed effect can be explained by a random variation. This supports the fact that we can only see (Fig. 2) different total incorporations of ^{14}C-acetate into tissue lipids, depending on the type of apical medium, when we add the precursor on the apical side. When the precursor is added on the basolateral side there is no difference to be seen with different apical media.

2. In addition to the significant interaction AB there is a significant general effect of A. To evaluate this significance we need to include the observation that there is *no* significant general effect of B: When we test B against the significant interaction AB by calculating $\mathrm{F} = s_3^2/s_2^2$ we find that the result is close to unity. This means that B alone is characterized by AB and, consequently, that s_B^2 must be zero, which gives us an opportunity to obtain an additional measure of the effect of AB with two extra degrees of freedom. By combining AB and B we get $s_2^2 = n\, s_{\mathrm{AB}}^2 + s_0^2 = 0.207$ with 4 degrees of freedom. This can then be used to test the general effect of A by calculating $\mathrm{F} = s_4^2/s_2^2$,which turns out to be significant at $P < 0.025$*. The combination of AB and B is an important detail. AB alone, with just 2 degrees of freedom, could not have corroborated the general effect of A at $P < 0.05$, which is the maximum value normally regarded as being "statistically significant". AB alone would only have shown $P < 0.10$.
   A significant main factor A supports the direct impression from Fig. 2 that the incorporation of ^{14}C-acetate from the basolateral side is much higher than from the apical side. Even with the same growth medium L15 on the apical side as on the basolateral side, the incorporation of ^{14}C-acetate into tissue lipids is 2–5 times higher from the basolateral side than from the apical side. The insert cell culture is obviously polarized.

3.3 Incorporation of ^{32}P-Phosphate

A two-dimensional ANOVA based on the logarithms of the ^{32}P counts shown in Fig. 2, but *without including the apical results* in batch II, shows the same significant AB interaction at the level $P < 0.025$ as in Table 2. We refrain from presenting the calculations.

In the following, Table 3 shows a three-dimensional ANOVA based on the logarithm of *all* the available ^{32}P-results. In addition to the same two factors A and B as in Table 2, we include a third factor, which we shall call Z, to emphasize its deviation from two other factors. Z covers an *unknown and undefined biological difference* between batch I and batch II. In this case the extra two duplicate values

*The 97.5% fractile (Hald 1952*a*) is 12 in connection with A, but only 7.3 in connection with AB. The reason is that the degrees of freedom are different in the numerators and the denominators of the corresponding F-values.

within batches are included to show variation within the sets. All the other set results are just single values.

Without further comments, we present (Hald 1952*b*) a partitioning of the standard variances that applies to a three-dimensional ANOVA, a generalization of the two-dimensional ANOVA presented in section 2. Besides the main factor Z, it also includes the interactions AZ, BZ, and ABZ. The three factors A, B, and Z are varied at the levels k, m, and p, respectively, with on the average n values in each individual set. The relevant equations now become:

$$
\begin{aligned}
s_8^2 &\equiv \frac{\mathrm{SSD_A}}{k-1} = n\,mp\,s_{\mathrm{A}}^2 + n\,p\,s_{\mathrm{AB}}^2 + nm\,s_{\mathrm{AZ}}^2 + n\,s_{\mathrm{ABZ}}^2 + s_0^2, \\
s_7^2 &\equiv \frac{\mathrm{SSD_B}}{m-1} = n\,k\,p\,s_{\mathrm{B}}^2 + n\,p\,s_{\mathrm{AB}}^2 + n\,k\,s_{\mathrm{BZ}}^2 + n\,s_{\mathrm{ABZ}}^2 + s_0^2, \\
s_6^2 &\equiv \frac{\mathrm{SSD_Z}}{p-1} = n\,m\,k\,s_{\mathrm{Z}}^2 + n\,m\,s_{\mathrm{AZ}}^2 + n\,k\,s_{\mathrm{BZ}}^2 + n\,s_{\mathrm{ABZ}}^2 + s_0^2, \\
s_5^2 &\equiv \frac{\mathrm{SSD_{AB}}}{(k-1)\times(m-1)} = n\,p\,s_{\mathrm{AB}}^2 + n\,s_{\mathrm{ABZ}}^2 + s_0^2, \\
s_4^2 &\equiv \frac{\mathrm{SSD_{AZ}}}{(k-1)\times(p-1)} = n\,m\,s_{\mathrm{AZ}}^2 + n\,s_{\mathrm{ABZ}}^2 + s_0^2, \\
s_3^2 &\equiv \frac{\mathrm{SSD_{BZ}}}{(m-1)\times(p-1)} = n\,k\,s_{\mathrm{BZ}}^2 + n\,s_{\mathrm{ABZ}}^2 + s_0^2, \\
s_2^2 &\equiv \frac{\mathrm{SSD_{ABZ}}}{(k-1)\times(m-1)\times(p-1)} = n\,s_{\mathrm{ABZ}}^2 + s_0^2, \\
s_1^2 &\equiv \frac{\mathrm{SSD_0}}{k\times m\times p\times(n-1)} = s_0^2.
\end{aligned}
$$

Table 3 shows the same significant interaction AB as in Table 2. In contrast to the 6 degrees of freedom that characterize the residual s_0^2 in Table 2, we have only 2 degrees of freedom in Table 3. However, 4 more degrees of freedom can be obtained from the non-significant interactions ABZ and BZ.

We start at the bottom of Table 3 with the triple interaction ABZ in relation to the residual variation $s_1^2 = s_0^2$. We calculate s_2^2/s_1^2 and get an F-value equal to unity. This means that s_{ABZ}^2 must be zero and that s_2^2 can be used as an extra measure of s_0^2. By combining ABZ and the residual we thus get a strengthened $s_0^2 = 0.027$ with 4 degrees of freedom.

The new s_0^2 is then used to test the interaction BZ. We calculate s_3^2/s_0^2 and get, once more, an F-value not above unity. This, in turn, means that also s_{BZ}^2 must be zero and that we can combine BZ, ABZ, and the residual and get $s_0^2 = 0.021$ with 6 degrees of freedom as presented in Table 3. *Even though we only have 2 duplicate results in Table 3, we have nevertheless managed to achieve the same reliability of s_0^2 as in Table 2, which was based on 6 duplicate results.* It is the degrees of freedom that count, not just the number of duplicate values.

Continuing our way up in Table 3 we see that $\mathrm{F} = s_4^2/s_0^2$ shows a highly significant effect at the level $P < 0.0005$ which means that we can confirm an interaction AZ. Similarly, calculating $\mathrm{F} = s_5^2/s_0^2$, we get a significant interaction AB at the level $P < 0.001$. This is actually what we have been looking for all along, now confirmed at $P < 0.001$, while it was confirmed at only $P < 0.03$ in Table 2. We

Table 3. ANOVA of ^{32}P-incorporation from Fig. 2. Logarithmic values.

i	Variation	SSD	Df	s_i^2	F-value	Fractile
8	A: Between precursor additions	8.868	1	8.868	$\frac{8.868}{5.025} = 1.8$	
7	B: Between apical media	1.602	2	0.801	$\frac{0.801}{1.056} = 1$	
6	Z: Between batches	4.131	1	4.131	$\frac{4.131}{4.103} = 1$	
5	Interaction AB	2.112	2	1.056	$\frac{1.056}{0.021} = 50$	99.9%=27 ($P < 0.001$)
4	Interaction AZ	4.103	1	4.103	$\frac{4.103}{0.021} = 195$	99.95%=46 ($P < 0.0005$)
3	Interaction BZ	0.017	2	0.009	$\frac{0.009}{0.027} = 1$	
2	Interaction ABZ	0.055	2	0.027	$\frac{0.027}{0.026} = 1$	
1	Residual (within sets)	0.053	2	0.026		
	Total	20.941	13			
	Residual+ABZ	0.108	4	0.027		
	Residual+ABZ + BZ	0.125	6	0.021		
	AB + B	3.714	4	0.929		
	AZ + Z	8.234	2	4.117		

have shown once more that the effect of changing the apical medium depends on to which side of the cell culture we add the precursor.

Regarding the observed—highly significant—interaction AZ (Table 3), it is the only confirmed effect that involves the unknown difference between the two batches. We have shown that s^2_{BZ} is zero, and when we test the main factor Z against the significant interaction AZ by calculating s_6^2/s_4^2, the F-value is unity. This means that s^2_{Z} is also zero. In other words, when something "goes wrong" in our cell cultures (the factor Z) it just affects the total incorporation of ^{32}P-phosphate from the apical side, independent of the apical medium.

Finally, we have the two original main factors A and B. The latter needs to be tested against the significant AB, and when we calculate $\mathrm{F} = s_7^2/s_5^2$ we again get a neat value of unity, so that s^2_{B} must be zero. Testing A is more difficult because it must be tested against both of the two significant interactions AB and AZ, involving the sum $s_4^2 + s_5^2$, corrected for an extra s_0^2. But what should the degrees of freedom in the denominator be? Since the numbers in Table 3 only hint at an F-value a little larger than unity, we leave it at that and postulate that if there is a general effect of A, as in Table 2, it is very small.

3.4 Incorporation of ^{14}C-Acetate (Three-Dimensional ANOVA)

What would it look like if we presented the ^{14}C-values in Fig. 2 in a three-dimensional ANOVA, which included the difference between batch I and batch II as a factor Z, just as in Table 3? The result is presented in Table 4 without further explanation.

Compared to the equivalent two-dimensional ANOVA in Table 2, Table 4 shows the same statistically significant main effect A and interaction AB. We note that AB in Table 4 is confirmed at the level $P < 0.001$—just as in Table 3—compared to AB at $P < 0.03$ in Table 2. This is because Table 4 (and Table 3) test against a s_0^2 "within batches", while Table 2 tests against a s_0^2 "between batches". Obviously, the latter is larger than the former.

Table 4 presents a significant main effect Z at the level $P < 0.005$. Note that AZ is zero, in contrast to Table 3, where AZ was highly significant and Z was zero. In Table 4 the unknown and undefined difference between batch I and batch II, regarding the incorporation of ^{14}C-acetate into tissue lipids, is shown as a simple enhancement of a general variation, similar to the random component s_0^2. In contrast to the equivalent variation, regarding incorporation of ^{32}P-phosphate, shown in Table 3, the batch-to-batch variation in Table 4 is not dependent on, whether the precursor is added on the apical or the basolateral side. This stresses the metabolic difference between the incorporation of ^{14}C-acetate and ^{32}P-phosphate. It justifies that s_Z^2 is contained in s_0^2 in Table 2

The main effect A is confirmed at $P < 0.025$ in both Table 2 and Table 4. It is tested against the significant interaction AB, i.e. $F = s_4^2/s_2^2$ in Table 2 and $F = s_8^2/s_5^2$ in Table 4, and is thereby practically independent of the whether or not s_0^2 also contains s_Z^2.

4 Discussion

Above all, we want to stress the merits of an analysis of variance. It is a very useful statistical tool that can be utilized by any experimentalist who cares to make the effort. An analysis of variance applied within radioecology and clinical medicine has previously been presented (Hansen 1994, Hansen & Westengaard 1966).

In the present case it has been a matter of isolating a troublesome extra variation. We have managed to extract the random component s_0^2 from a batch-to-batch variation, which also included an unknown systematic factor. Thereby we were able to test the two induced factors—here apical or basolateral precursor addition and different apical media—against a common random component s_0^2 within each of the two batches. By regarding the unknown systematic factor (designated Z) at a level equal to that of the two induced factors, A and B, in a three-dimensional analysis of variance, we have also to some degree been able to characterize what "went wrong" between the two batches.

Batch-to-batch variations between cell cultures and year-to-year variations between animal cohorts are more common than most of us like to admit. Based on the Student's t-test, a significant year-to-year variation of the response to radiation-induced stress in eels, captured from the wild, has previously been reported (Hansen 1991). The investigation concluded that because an effect is not

Table 4. ANOVA of ^{14}C-incorporation from Fig. 2. Logarithmic values.

i	Variation	SSD	Df	s_i^2	F-value	Fractile
8	A: Between precursor additions	24.540	1	24.540	$\frac{24.540}{1.332} = 18$	97.5%=12 ($P < 0.025$)
7	B: Between apical media	2.711	2	1.356	$\frac{1.356}{1.309} = 1$	
6	Z: Between batches	0.973	1	0.973	$\frac{0.973}{0.039} = 25$	99.5%=16 ($P < 0.005$)
5	Interaction AB	2.618	2	1.309	$\frac{1.309}{0.039} = 34$	99.9%=22 ($P < 0.001$)
4	Interaction AZ	0.058	1	0.058	$\frac{0.058}{0.053} = 1$	
3	Interaction BZ	0.002	2	0.001	$\frac{0.001}{0.053} = 1$	
2	Interaction ABZ	0.043	2	0.022	$\frac{0.022}{0.084} = 1$	
1	Residual (within sets)	0.167	2	0.084		
	Total	31.112	13			
	Residual+ABZ	0.210	4	0.053		
	Residual+ABZ + BZ + AZ	0.270	7	0.039		
	AB + B	5.329	4	1.332		

easily reproducible, it does not mean that it should not be taken into account.

We note that Fig. 2 only shows a few error bars. This is not disturbing when we regard the ^{14}C-results, since these in any case are evaluated in Table 2 as pairs. The ^{32}P-results are different. Here we seem to break all the rules, when we base our conclusions in Table 3 on single assays. At first there is much to be said to support such an objection: The observed ^{32}P-results are actually out of full biological control. But that is besides the point. It is more reasonable to stress that the systematic nature of an analysis of variance makes it possible to compensate for the lack of reproducible duplicate values. When we, for example in Table 3, see BZ as zero, which means that the pattern of response to changes in the apical medium (the factor B) is the same in both batches—relative to FW, with apical addition of ^{32}P-phosphate, results are divided by two in FW+M and multiplied by three in L15, cf. Fig. 2—this is a very powerful indication of sufficient reproducibility to enable a further evaluation of other effects such as for example the interaction AB. As already mentioned above, it is the number of degrees of freedom that counts, independent of its origin. Full biological control is not necessary and often impossible.

The significant interaction AB—i.e. that the effect of changing the apical medium depends to which side of the tissue we add the radioactive precursor—is common to both the incorporation into tissue lipids of ^{14}C-acetate and ^{32}P-phosphate. This is the major confirmed effect in the present investigation. Together with the significant main factor A—apical or basolateral addition of precursor— in connection with just the ^{14}C-results, it stresses the polarization of branchial tissue

grown in insert cups. Similar to gill tissue in vivo, the "top" (apical) membrane of coherent branchial cells, facing the inner volume of the insert cup, is clearly seen to represent a different type of boundary layer than the "bottom" (basolateral) membrane facing the permeable bottom of the insert cup.

We could not in Fig. 2 observe any closer resemblance between FW+M and L15 regarding the apical incorporation of radio-labeled precursors: Addition of mannitol to FW in an amount equivalent to the osmolality of L15 was not enough to mimic the stimulating effect of L15 relative to FW. This answers one of the original questions in this investigation. It appears directly from Fig. 2, without involving an analysis of variance.

We note that, although the two-dimensional ANOVA in Table 2 gives sufficient information regarding the incorporation of ^{14}C-acetate, the introduction of the factor Z in Table 4 nevertheless tells us something more. It confirms an enhanced, apparently random, variation between batches relative to s_0^2 within batches. There is no harm done by introducing an unnecessary extra factor. Since it is just a matter of which computer-program option is applied, Table 4 might in any case replace Table 2, notwithstanding that is more complicated.

Finally, concerning the significant interaction AZ, it again refers to tissue polarization. The difference between batch I and batch II in Table 3 has only to do with the incorporation of ^{32}P-phosphate from the top, apical side. As suggested previously (Hansen et al. 2002), the apical membrane seems to be the main site of a lipid-induced regulation of ion transport. Here we have additional information. We could also show that while the incorporation in batch I lead to the previously seen ^{32}P-phospholipid pattern in FW—dominated by ^{32}P-phosphatidylethanolamine (^{32}P(PE)) (Hansen et al. 2002)—this was not the case in batch II, where all the ^{32}P-phospholipid patterns were dominated by ^{32}P-phosphatidylcholine; i.e it was batch II that seemed to be the "wrong" one. Moreover, the fact that (cf. Fig. 2) an extra low apical incorporation of ^{32}P-phosphate in batch II was accompanied by an even higher apical incorporation of ^{14}C-acetate in batch II than in batch I, agrees with previous results in vivo (Hansen et al. 1992), namely that low $^{32}\mathrm{P}/^{14}\mathrm{C}$ incorporation ratios normally result in a low relative incorporation into ^{32}P(PE). This means that what was wrong in batch II could seem to be part of an otherwise normally seen process. Batch II could perhaps just have been at a premature stage relative to batch I. This needs a further investigation.

5 Conclusion

We have argued in favor of evaluating troublesome batch-to-batch variations rather than disregarding them. The issue is also that of scientific ethics. Is it reasonable to reject genuine experimental results, when we don't know a priori that something went wrong? Formally, the purpose of any scientific investigation is to disprove a theory by checking its consequences (Popper 1963). Not before we have failed to disprove the proposed theory many times can we, in principle, say that there is circumstantial evidence for its correctness. From this point of view one should not conceal "awkward" experimental results.

This stern admonition should not prevent us to emphasize a positive aspect of our procedure: A general effect—here the interaction AB—that can be confirmed in spite of partially irreproducible results, stands out as additionally strengthened in a final evaluation. A factorial experimental design does not necessarily require an a

priori theory. It is possible to follow Popper (1963) and yet retain a fundamentally positive approach. An ANOVA is in itself strictly neutral. It actually ends up by estimating the possibility that an observed difference can be negated.

Acknowledgements

The present investigation i based on results closely connected with previous work (Hansen et al. 2002) in collaboration with Scott P. Kelly and Chris M. Wood, McMaster University, Ontario, Canada. Claus B. Andersen of Risø National Laboratory was very helpful with statistical calculations and figure preparation. We appreciate the helpful support from Benny Majborn and Sven Nielsen.

References

Bailey, N. T. J. (1958), *Statistical Methods in Biology*, English University Press, London.

Bøtter-Jensen, L., Hansen, H. J. M. & Theodorsson, P. (1977), 'A multicounter system for scanning ultra-low-level radio-chromatograms', *Nucl. Instrum. Methods* **144**, 529–632.

Dytham, C. (1999), *Choosing and Using Statistics: a Biologist's Guide*, Blackwell Science, Oxford.

Gilmour, K. M., Pärt, P., Pisam, M., McDonald, D. G. & Wood, C. M. (1998), 'Permeability and morphology of a cultured epithelium from the rainbow trout during prolonged apical exposure to freshwater', *J. Exp. Zool.* **281**, 531–545.

Hald, A. (1952*a*), *Statistical Tables and Formulas*, John Wiley & Sons, Inc., New York.

Hald, A. (1952*b*), *Statistical Theory with Engineering Applications*, John Wiley & Sons, Inc., New York.

Hansen, H. J. M. (1975), 'Changes in the lipid fraction of eel gills after ionizing irradiation *in vivo* and a shift from fresh to sea water', *Radiat. Res.* **62**, 216–224.

Hansen, H. J. M. (1991), 'Significant year-to-year variation of the response to radiation-induced stress shown by gill fatty acid metabolism in eels (Anguilla anguilla) captured from the wild', *Comp. Biochem. Physiol.* **100C**, 99–101.

Hansen, H. J. M. (1994), Statistical treatment of data, *in* 'Radioecology', World Scientific, Singapore, pp. 329–337.

Hansen, H. J. M., Kelly, S. P., Grosell, M. & Wood, C. M. (2002), 'Studies on lipid metabolism in trout (Oncorhynchus mykiss) branchial cultures', *J. Exp. Zool.* **293**, 683–692.

Hansen, H. J. M., Olsen, A. G. & Rosenkilde, P. (1992), 'Comparative studies of lipid metabolism in salt-transporting organs of the rainbow trout (Oncorhynchus mykiss). Further evidence of mono-unsaturated phosphatidylethanolamine as a key substance', *Comp. Biochem. Physiol.* **103B**, 81–87.

Hansen, H. J. M. & Westengaard, E. (1966), 'Et klinisk faktorforsøg (A clinical factorial experiment)', *Ugeskr. Læg.* **128**, 465–470. In Danish.

Popper, K. R. (1963), *Conjectures and Refutations*, Routledge, London.

Wood, C. M. & Pärt, P. (1997), 'Cultured branchial epithelia from freshwater gills', *J. Exp. Biol.* **200**, 1047–1059.

Bibliographic Data Sheet **Risø–R–1384(EN)**

Title and author(s)

Batch Variation between Branchial Cell Cultures: An Analysis of Variance

Heinz J. M. Hansen, Martin Grosell & Leif Kristensen

ISBN	ISSN
87-550-3163-3; 87-550-3161-7(internet)	0106-2840

Dept. or group	Date
Radiation Research Department	January 31, 2003

Groups own reg. number(s) Project/contract No.

Sponsorship

Pages	Tables	Illustrations	References
20	4	2	14

Abstract (Max. 2000 char.)

We present in detail how a statistical analysis of variance (ANOVA) is used to sort out the effect of an unexpected batch-to-batch variation between cell cultures. Two separate cultures of rainbow trout branchial cells were grown on permeable filter supports ("inserts"). They were supposed to be simple duplicates for testing the effect of two induced factors—apical or basolateral addition of radioactive precursors and different apical media–on the incorporation of ^{14}C-acetate and ^{32}P-phosphate into tissue lipids. Unfortunately, they did not altogether give the same result. By accepting this fact and introducing the observed difference between batches as one of the factors in an expanded three-dimensional ANOVA, we were able to overcome an otherwise crucial lack of sufficiently reproducible duplicate values. We could thereby show that the effect of changing the apical medium was much more marked when the radioactive lipid precursors were added on the apical, rather than on the basolateral, side. The insert cell cultures were obviously polarized. We argue that it is not reasonable to reject troublesome experimental results, when we do not know a priori that something went wrong. The ANOVA is a very useful statistical tool that can be utilized by any experimentalist who cares to make the effort.

Descriptors

APICAL;BASOLATERAL; FISH; GILLS; LIPID METABOLISM; MEMBRANE; ONCORHYNCHUS MYKISS; PHOSPHILIPIDS;POLARISATION; SALINITY; SCIENTIFIC ETHICS